



A Hierarchical Method for Locating the Interferometric Fringes of Celestial Sources in the Visibility Data

Rong Ma¹, Ruiqing Yan¹, Hanshuai Cui¹, Xiaochun Cheng², Jixia Li³, Fengquan Wu³, Zongyao Yin¹, Hao Wang¹,
Wenyi Zeng¹, and Xianchuan Yu¹

¹ School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China; yuxianchuan@163.com

² Department of Computer Science, Swansea University, Swansea, SA1 8EN, UK

³ National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

Received 2023 June 1; revised 2023 October 8; accepted 2023 October 30; published 2024 February 26

Abstract

In source detection in the Tianlai project, locating the interferometric fringe in visibility data accurately will influence downstream tasks drastically, such as physical parameter estimation and weak source exploration. Considering that traditional locating methods are time-consuming and supervised methods require a great quantity of expensive labeled data, in this paper, we first investigate characteristics of interferometric fringes in the simulation and real scenario separately, and integrate an almost parameter-free unsupervised clustering method and seeding filling or eraser algorithm to propose a hierarchical plug and play method to improve location accuracy. Then, we apply our method to locate single and multiple sources' interferometric fringes in simulation data. Next, we apply our method to real data taken from the Tianlai radio telescope array. Finally, we compare with unsupervised methods that are state of the art. These results show that our method has robustness in different scenarios and can improve location measurement accuracy effectively.

Key words: methods: data analysis – techniques: image processing – techniques: interferometric

1. Introduction

Dark energy (Korytov et al. 2019; Tanoglidis et al. 2021; Everett et al. 2022) detection is a very important topic in cosmology. As resolution and sensitivity of radio telescope technology become higher and higher, astronomical researchers can observe the universe in wider frequency ranges. The Tianlai project (Chen 2011, 2012) aims at exploring the large-scale structure by measuring the redshifted 21 cm emission line of neutral hydrogen. It is located at Hongliuxia Observing Station in the northeast of Xinjiang province of China. Tianlai records the signal of sources in the form of visibility data. One can obtain fringes directly from the raw visibility data by taking out either the amplitude or the phase part of the complex value. Hereafter, we treat the phase part of the visibility data within some time interval and 700–800 MHz in frequency band as one image, because the phase part is usually more sensitive than the amplitude. Through a series of calibration and data analysis activities, specific parameters of sources and a sky map can be recovered (Li et al. 2020; Zuo et al. 2021). Among them, one important step is locating interferometric fringes of sources in the raw visibility data. In the general case, these images include interferometric fringes in different low signal-to-noise ratios (S/Ns). Locating these interferometric fringes in images with high accuracy will decide the accuracy of downstream physical parameter estimation tasks. How to develop a high accuracy

measurement method for these weak signal processing tasks remains a challenge.

In past years, locating an interferometric fringe and physical parameter estimation depend on the technician and researchers' experience, which consume a lot of human labor but obtain low efficiency and accuracy. As artificial intelligence (AI) develops, more methods based on AI show remarkable accuracy in weak signal processing. On the other hand, with the help of high performance computers, running time is reduced drastically. Hence, we consider applying AI methods to locating an interferometric fringe in raw visibility data.

In recent years, AI methods provide effective tools to explore astronomical problems. Traditional machine learning methods and deep learning methods can tackle small and large scale scenarios. Cavaglia et al. (2018) proposed a method based on random forests and genetic programming for gravitational wave detection. Gheller et al. (2018) utilized a convolutional neural network to detect sources from extragalactic sources. Deep learning has achieved much progress in many fields such as image classification (He et al. 2016) and segmentation (Shelhamer et al. 2017), object detection (Redmon et al. 2016) and signal detection (Awni et al. 2019). Liu et al. (2019a, 2019b) introduced a deep convolutional neural network for large scale stellar spectra classification and detecting candidates of supernova remnants. Yan et al. (2022) introduced channel attention shrinkage networks into weak source fringe

detection. Wang et al. (2019) adopted ResNet to construct a system to select pulsar candidates. Furthermore, some researchers introduced transfer learning into different fields of study (Xu et al. 2015; He et al. 2022; Kuang et al. 2022). In real application, researchers developed more deep neural network architecture for multiple scenarios (Zeng et al. 2021; Fu et al. 2022; Lin et al. 2022; Peng et al. 2022).

However in most situations, supervised learning needs a large quantity of training data with labels and computation. As high performance computation devices develop, computational ability is not a big problem anymore, but the obstacle to obtain labeled data is still unresolved, especially in unsupervised and class imbalanced scenarios. Aiming at unsupervised industry measurement scenarios, new solutions are proposed in their fields (Cao et al. 2022; Fong & Narasimhan 2022; Zhu et al. 2022). Owing to expensive labeled data, in locating interferometric fringes, unsupervised methods are better choices for our tasks.

Fuzzy clustering is a powerful tool in real applications. Since Zadeh (1965) proposed a fuzzy set, many unsupervised methods were subsequently developed based on this theory. Exempli gratia, Dunn (1973) proposed the fuzzy C-means (FCM) algorithm to discard hard clustering and enhance clustering accuracy. Krinidis & Chatzis (2010) constructed an objective function with a local fuzzy factor G_{ki} to improve the accuracy. Zeng et al. (2020) introduced hesitant fuzzy theory and proposed the hesitant fuzzy C-means (HFCM) algorithm. Gong et al. (2013) merged a kernel method and local information through weights to propose kernel metric weighted fuzzy C-means algorithm with local information (KWFLICM). To generalize KWFLICM, Memon & Lee (2018) proposed neighbor searching methods based on KWFLICM to deal with high dimensional data. These effective fuzzy clustering variants have been applied in image segmentation tasks successfully.

In our task, although acquiring visibility data in a simulation scenario through programming is easy, labeled data from real scenarios are expensive due to labeling relying on expert experience. To overcome this disadvantage, we consider using unsupervised methods. Considering that massive uncertainty and noise exist in simulation and real visibility data, fuzzy clustering is more suitable. Comparing with other variants of FCM, the KWFLICM algorithm is a representative with robustness and satisfactory accuracy in segmentation tasks. Despite the KWFLICM algorithm providing a good segmentation result, other factors and noises still cause influences for downstream physical parameter estimation. To avoid these bad influences, we further select representative regions as signals. Therefore, in this paper, we first investigate the characteristics of the images in simulated and real scenarios, and propose a hierarchical method for locating an interferometric fringe. Aiming at distinguishing interferometric fringe signals from background, we utilize the KWFLICM algorithm to complete the image segmentation. Further, we use a seed filling

algorithm to remove the influence of noise and retain most interferometric fringe features for downstream task processing. We regard the width of maximum connection region as the width of an interferometric fringe. Hence, we propose a complete plug and play solution with high accuracy for locating interferometric fringes in raw visibility data.

The organization of this paper is as follows. In Section 2, we review a classical unsupervised clustering method and its advanced variant. In Section 3, we propose a novel hierarchical method for locating an interferometric fringe. In Section 4, experiments in simulated and real scenarios are provided to illustrate the effectiveness and validation of our method, meanwhile, we make a comparison with the state of the art unsupervised methods to illustrate the effectiveness of our method. Finally, the conclusion is drawn in Section 5.

2. Preliminaries

Since Dunn (1973) proposed the FCM algorithm, and Hathaway & Bezdek (2000) gave an extension, FCM has attracted many researchers' interests. It is well known that fuzzy clustering provides an effective way to segment an image. Comparing with deep learning methods, FCM and its effective variants can produce segmentation results in an iterative way. In many scenarios with limited samples, unsupervised segmentation methods are more efficient. In image segmentation tasks, an image will be segmented into different regions according to different features, such as gray level, local information, texture and so on. For the classical FCM algorithm, it constructs an objective function J_m through the sum of the squared error function in Equation (1)

$$J_m = \sum_{i=1}^N \sum_{j=1}^c u_{ji}^m d_{ji}, \quad (1)$$

where $X = \{x_1, x_2, \dots, x_N\}$ represents the data set consisting of N data samples. In an image segmentation situation, every sample has only one-dimension, but in other situations, each sample can have even higher dimensions. The quantity c ($c \in [1, N]$) represents the number of clusters, which is c different regions in the segmentation. u_{ji} signifies the membership degree and describes the degree of sample x_i belonging to cluster v_j . m is an exponent factor that ensures the algorithm converges to an optimal value. d_{ji} describes the distance between pixel x_i and cluster center v_j . The termination of the algorithm is controlled by the threshold ϵ , when $\epsilon > \sum |V^{(b+1)} - V^{(b)}|$.

Objective function J_m can be solved through the Lagrange multiplier method, and we can obtain a segmentation result through Equations (2) and (3). In these equations, $u_{ji}^{(b)}$ represents membership from the b th iteration. $v_j^{(b+1)}$ refers to

clusters from the $(b + 1)$ th iteration.

$$v_j^{(b+1)} = \frac{\sum_{i=1}^N (u_{ji}^{(b)})^m x_i}{\sum_{i=1}^N (u_{ji}^{(b)})^m}, \quad (2)$$

$$u_{ji}^{(b+1)} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ji}}{d_{ki}} \right)^{\frac{2}{m-1}}}. \quad (3)$$

Gong et al. (2013) merged the kernel method and local information carried by neighboring pixels, which enhanced the ability to handle outliers and improved the robustness of FLICM. In this algorithm, a novel objective function with local similarity factor G'_{ki} is constructed by

$$J_m = \sum_{i=1}^N \sum_{k=1}^c u_{ki}^m (1 - K(x_i, v_k)) + G'_{ki}, \quad (4)$$

where the novel factor G'_{ki} can be described as follows

$$G'_{ki} = \sum_{i=1}^N \sum_{k=1}^c u_{ki}^m \sum_{i \neq j, j \in N_i} w_{ij} (1 - u_{ki})^m (1 - K(x_j, v_k)). \quad (5)$$

Here w_{ij} is called the fuzzy factor which describes the weight between the center pixel i and neighbor pixel j . $K(x_j, v_k)$ represents the distance based on a kernel function. $(1 - u_{ki})^m$ is a penalty term, which can accelerate the convergence speed of the algorithm.

3. Methodology

In this section, we investigate the characteristics of images with a fringe and introduce the KWFLICM algorithm to process these images. We regard this task as an image segmentation task in the first stage. After obtaining segmentation results, we further process them with a seed filling algorithm to locate the interferometric fringe in the second stage. Especially, for a scenario with multiple interferometric fringes, we propose an eraser algorithm to iteratively find all fringe locations in the second stage. Generally speaking, our method has two stages and we will introduce our method in detail.

3.1. Characteristics of Interferometric Fringe in Simulated and Real Scenarios

Images with an interferometric fringe have differences in simulated and real scenarios. In a simulation scenario, the interferometric fringe is obvious and has little noise, so it is easier to distinguish the location of an interferometric fringe compared with real scenarios under the same S/N. From intuition, locating an interferometric fringe is harder when the S/N values are lower, and as S/N becomes higher, the shape of the interferometric fringe is more explicit. In a real scenario, interferometric fringe signals and noises are usually interlaced.

In some special situations, an interferometric fringe with low S/N will disappear in noises.

3.2. Kernel Metric Weighted Fuzzy C-means Algorithm with Local Information

For different interferometric fringes with different S/Ns, we adopt an unsupervised fuzzy clustering method. Its advantages are that we can obtain a segmentation result of a given data with low computational cost. The FCM algorithm is a popular clustering algorithm. Since it was first proposed by Dunn (1973), many researchers have developed many variants of the FCM algorithm, and the KWFLICM algorithm is a representative. This algorithm fuses the kernel method and local information through weights to obtain good clustering effectiveness.

Sometimes, data are hard to handle in low dimensions. Therefore, projecting low dimensional data into higher dimensions is a good solution. However, computational cost for high dimensional data is very expensive. To solve this problem, the kernel method is introduced to reduce computation. In traditional fuzzy clustering algorithms applied in image segmentation, many algorithms do not incorporate local information and the segmentation results are not good enough. Relations between pixels are a key factor that can influence the final segmentation result. Neighboring pixels usually carry similar information with the central pixel. Therefore, the KWFLICM algorithm, incorporating local information and the kernel method, provides us with higher accuracy.

From Equation (4), the partial derivatives of J_m with respect to membership degree u_{ki} and cluster centers v_k are computed separately. The update equations for clusters (Equation (6)) and membership degree (Equation (7)) can be obtained by setting these partial derivatives to be 0.

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m K(x_i, v_k) x_i}{\sum_{i=1}^N u_{ki}^m K(x_i, v_k)}, \quad (6)$$

$$u_{ki} = \left(\sum_{l=1}^c \left(\frac{1 - K(x_i, v_k) + \sum_{i \neq j, j \in N_i} w_{ij} (1 - u_{kj})^m (1 - K(x_j, v_k))}{1 - K(x_i, v_l) + \sum_{i \neq j, j \in N_i} w_{ij} (1 - u_{lj})^m (1 - K(x_j, v_l))} \right)^{\frac{1}{m-1}} \right)^{-1}. \quad (7)$$

Next, we can use this algorithm to process images with an interferometric fringe. For a color image, it always has the three channels of R, G, B. Here, we can transform it into a gray level image, which helps in reducing the run time. Then, the algorithm will initialize some important parameters, and produce a membership degree matrix and centers matrix according to the image pixels. The algorithm will continue iterations until convergence by using Equations (6) and (7). When the algorithm terminates, the new image has only two regions, which represent the signals and noises (see the proof in

Gong et al. 2013). We conclude the details of KWFLICM in Algorithm 1.

Algorithm 1. KWFLICM algorithm—First Stage of Our Method

Input:

Gray level image with interferometric fringe.

Output:

Image segmented by two regions.

Steps:

- 1: Set parameters: number of clusters c , exponent fuzzy factor m , window size N and stopping threshold ϵ .
 - 2: Randomly initialize c clusters and membership degree matrix U .
 - 3: Set counter $b = 0$.
 - 4: Compute fuzzy weight factor w_{ij} and update distance $K(x_j, v_k)$.
 - 5: Update centers matrix $V^{(b+1)}$ by using Equation (6).
 - 6: Update membership degree matrix $U^{(b+1)}$ by using Equation (7).
 - 7: If $|\sum V^{(b+1)} - V^{(b)}| < \epsilon$, then terminate the algorithm, otherwise, set $b = b + 1$ and go to Step 4.
-

3.3. Maximum Connection Region Finding Algorithm

Segmentation result retains features of an interferometric fringe. To extract the most important feature, we adopt the seed filling algorithm to obtain the maximum connection region. For result in the above step, although signals and noise have been divided, noise regions in the image still have influences. We compute the area of every enclosed region Reg_s , ($s = 1, 2, \dots, m$) through counting pixels in the segmentation result, where m represents number of connection regions. Some regions with a large area are usually signal and contain important features. Others with a small area are usually noise. Further, we sort these regions according to their area in descending order and choose the maximum connection region as output of the algorithm.

We list the computing steps of the seed filling algorithm in Algorithm 2.

Algorithm 2. Seed Filling Algorithm—Second Stage of Our Method

Input:

Output image of Algorithm 1.

Output:

Maximum connection region of segmentation result.

Steps:

- 1: Input image segmented from KWFLICM algorithm.
 - 2: Compute different areas of connection regions Reg_s .
 - 3: Sort all connection regions according to area, and choose region with the biggest area $\text{MAX}(\text{Reg}_s)$.
 - 4: Output image with maximum connection region $\text{Reg}_s^{\text{MAX}}$.
-

3.4. Eraser Algorithm

In special situations, an image can include multiple interferometric fringes. To deal with this problem, we propose the eraser algorithm in Algorithm 3. Generally speaking, the eraser algorithm will iteratively run. In every iteration, the algorithm finds and retains a maximum connection region, and erases all relative regions in the width range of maximum connection region until obtaining a given number of maximum connection regions.

In this algorithm, *num* is used to control termination of the algorithm. To explore this parameter, we count signal pixels by column and make a statistical curve. After curve smoothing and peak finding, we can regard the number of peaks as the value of parameter *num*.

Algorithm 3. Eraser algorithm—For multiple interferometric fringe scenario

Input:

Output image of Algorithm 1, the number of interferometric fringes *num*, iteration counter *counter*=0.

Output:

num maximum connection region location.

Steps:

- 1: Input segmentation result from KWFLICM algorithm.
 - 2: Find and retain maximum connection region through Algorithm 2.
 - 3: *counter* = *counter* + 1.
 - 4: If *counter* < *num*, erase connection regions within width range, go to step 2, else go to step 5.
 - 5: Terminate algorithm and compute location of different fringes.
-

Example: To illustrate this process, we make this example and diagram in Figure 1. In the figures, the images are the phase part of the raw visibility. The horizontal axis is time and vertical is frequency with origin in the bottom left corner. For simplicity, we have omitted the coordinates. The image size is randomly set to be 224×224 . Assuming we know that there are three interferometric fringes, our algorithm will complete three iterations to find three maximum connection regions. We find the first maximum connection region, and retain this region. Then we obtain the width of this region ((166,0), (190,223)) and erase all regions in this width range (connection regions in red box). Then we find the maximum connection region ((23,0), (62,223)) in residual regions, and erase all regions in this width range again. Finally, we find the maximum connection region ((99,0), (127,223)) in the residual regions again. Through three iterations, we obtain three maximum connection regions, and the algorithm reaches termination conditions. This algorithm can help to generalize the range of application of our method. When image includes

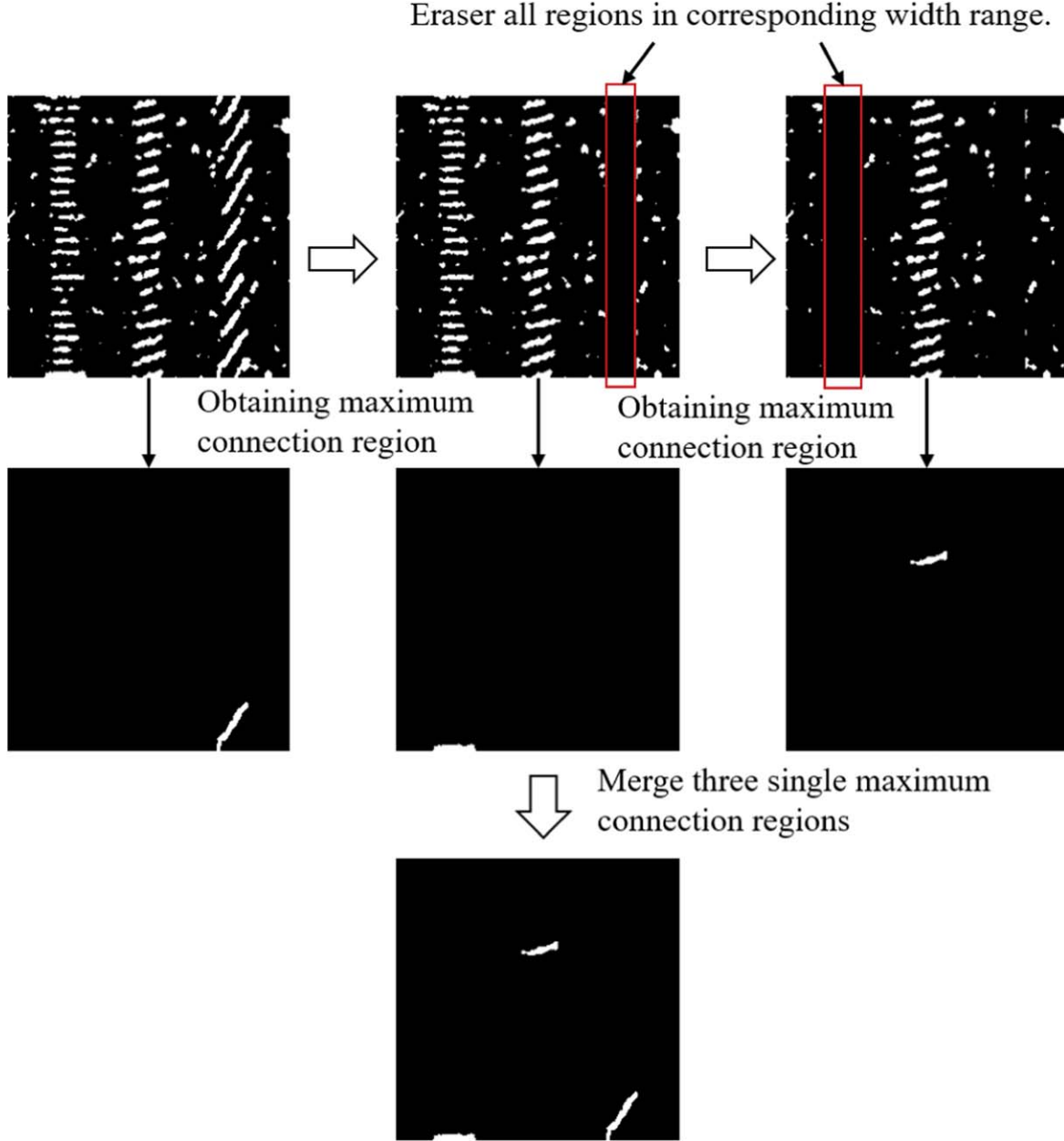


Figure 1. Eraser algorithm process. The images are the phase part of the raw visibility. The horizontal axis is time and vertical is frequency with origin in the bottom left corner. The image size is randomly set to be 224×224 . Three fringes are simulated, so three iterations are run to locate the maximum connection regions. In each iteration, the maximum connection region is found and then erased. Finally, all three maximum connection regions $[(166,0), (190,223)]$, $[(23,0), (62,223)]$, $[(99,0), (127,223)]$ are merged into the same image.

only one interferometric fringe, the eraser algorithm will degenerate to the classical seed filling algorithm.

We integrate these algorithms above to construct our hierarchical method. To illustrate the process of this hierarchical method intuitively, the flowchart is shown in Figure 2.

4. Experiments

In this section, we display how to produce simulation data, and provide a validation of our method in different scenarios and a comparison with state of the art unsupervised methods.

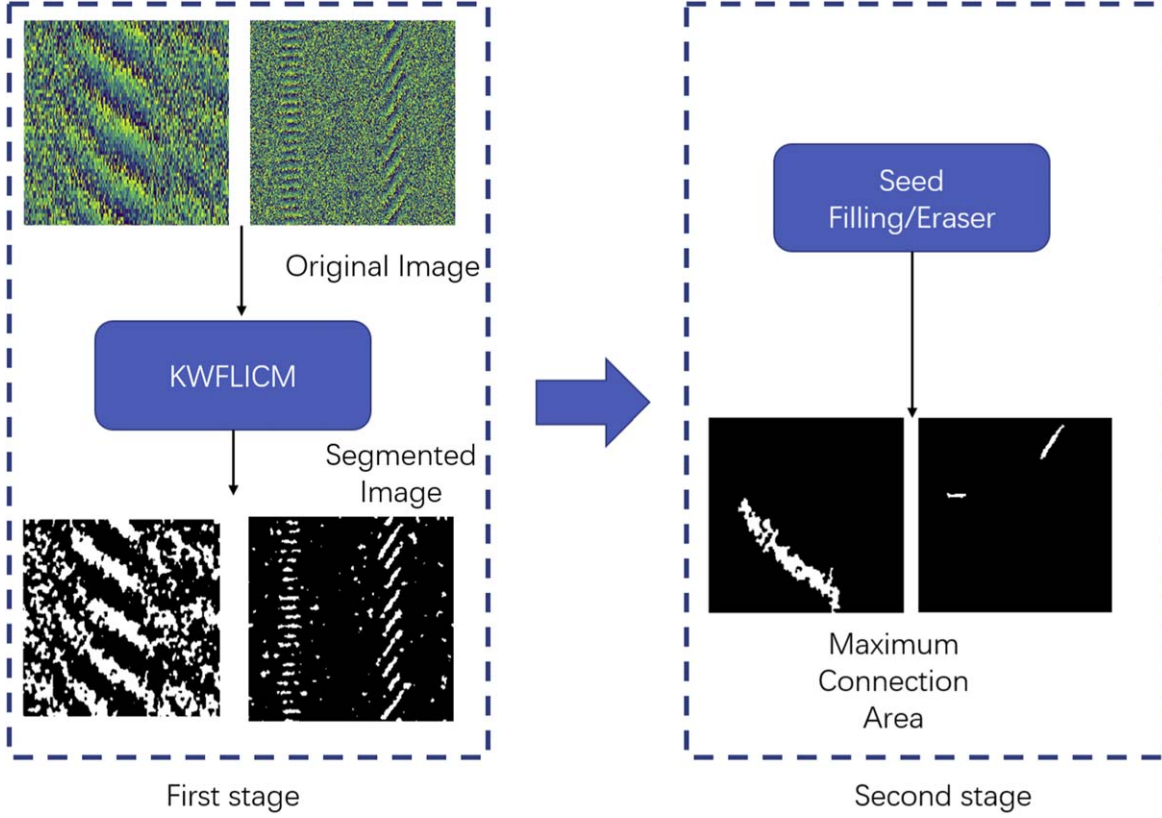


Figure 2. Flowchart of our method for locating the interferometric fringe of a source in the visibility data. This method can be divided into two stages: the first stage can be regarded as the image segmentation task, and the second stage can be used to distinguish the signal and suppress the noise.

Table 1

Baseline, Source Location, Fringe Length and S/N of Simulation Images

Simulation Image	Baseline Length (m)	Source Location (%)	Fringe Length (%)	S/N
(a)	11	0	15	0.9
(b)	17	100	15	0.8
(c)	23	80	15	0.5
(d)	28	70	15	0.6
(e)	32	20	15	1.3
(f)	36	40	15	0.7
(g)	42	30	15	1.5
(h)	47	50	15	1.0
(i)	49	10	15	1.1

4.1. Simulation Data

The visibility for a pair of antennas can be expressed as below (Thompson et al. 1991),

$$V = A \exp\left(j \cdot 2\pi f \cdot \frac{\mathbf{B} \cdot \mathbf{n}_s}{c}\right) + N, \quad (8)$$

where j is the imaginary unit, f is the frequency, \mathbf{B} is the baseline vector and \mathbf{n}_s is the direction vector from the antenna to the radio

source. The amplitude of the visibility A is the signal strength received by the interferometer when a point source transits over the antenna beam, thus reflecting the beam shape. In the simulation scenario, the amplitude is commonly regarded as a Gaussian function. The peak value of the Gaussian curve is proportional to the brightness of the source. Besides the pure visibility generated by the source, the interferometer system inevitably receives instrumental noises. Here in our simulation, we simulate the noises with a normal distribution which is added to the visibility as N in Equation (8). Nevertheless, the instrumental noise can be rather complicated in the real scenario. This includes the cross couplings between adjacent feeds, the variation of the system gain induced by the varying temperature and even transient radio interference from nearby human activities. The S/N is determined by the ratio of the source's peak value and the standard deviation of the instrumental noise.

Hence, we complete computer programming according to Equation (8) above to produce simulation data and labels as pairs. We obtain results from our method and compare the results with labels to compute the accuracy. On the other hand, in the real scenario, we use real visibility data from the Tianlai telescope array to validate our method. Owing to real data not having labels, we employ experts to complete manual labels.

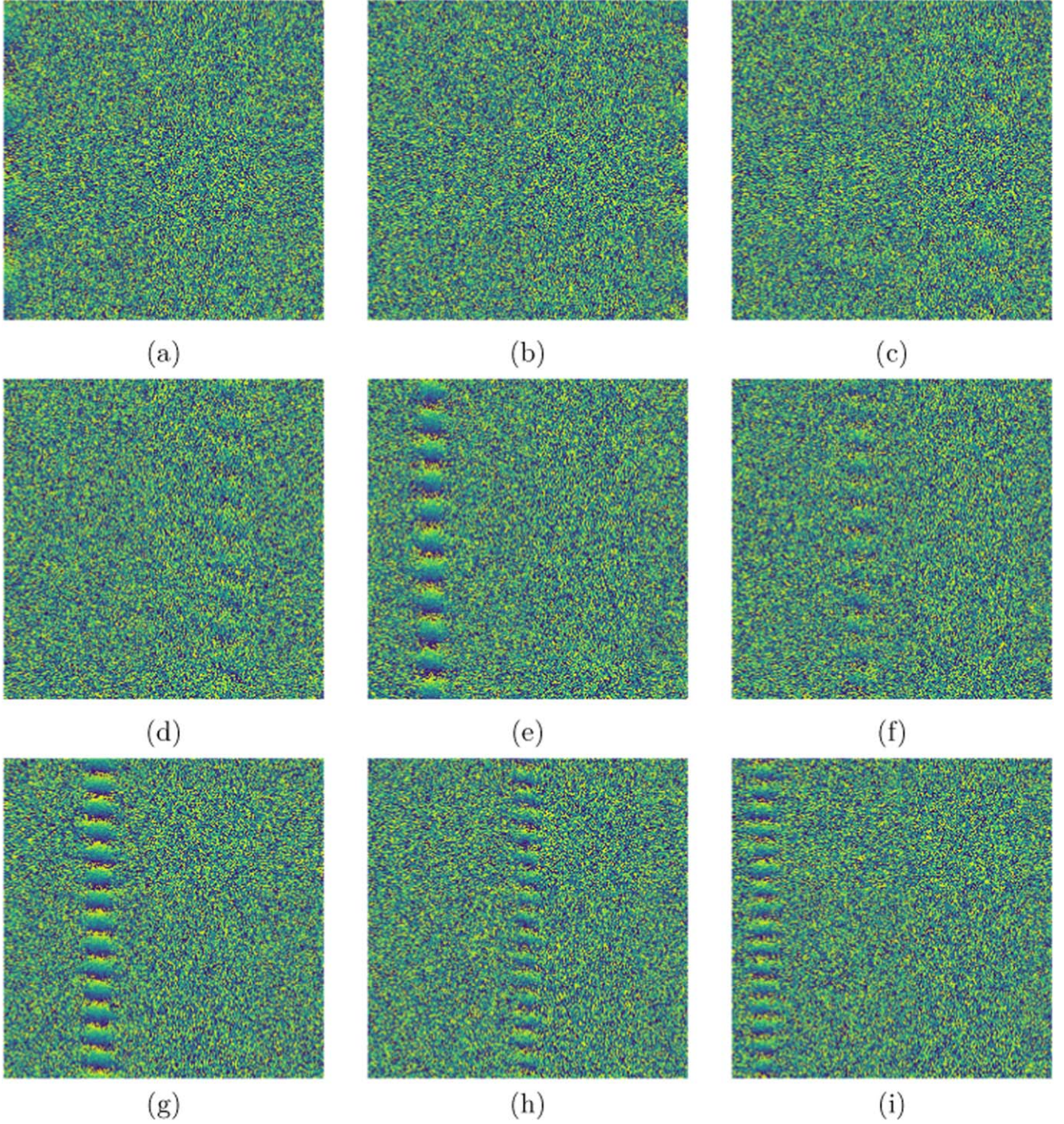


Figure 3. To validate our method, we randomly produce images with interferometric fringes in three main indexes: S/N, baseline length and source location through computer programming. Some images have explicit fringes, such as (g) and (e), and some have dim fringes, such as (c) and (d). We list more details on these images in Table 1.

4.2. Simulation Scenario

4.2.1. Single Interferometric Fringe

In this experiment, we produce the visibility images and list the details of these images in Table 1. The column Baseline

Length represents the distance between two antennas, and its range varies between 10 and 50 m. The column Source Location represents the location percentage relative to the whole 224 time points. The column Fringe Length is the ratio of the width of the source's fringe to the whole 224 time points.

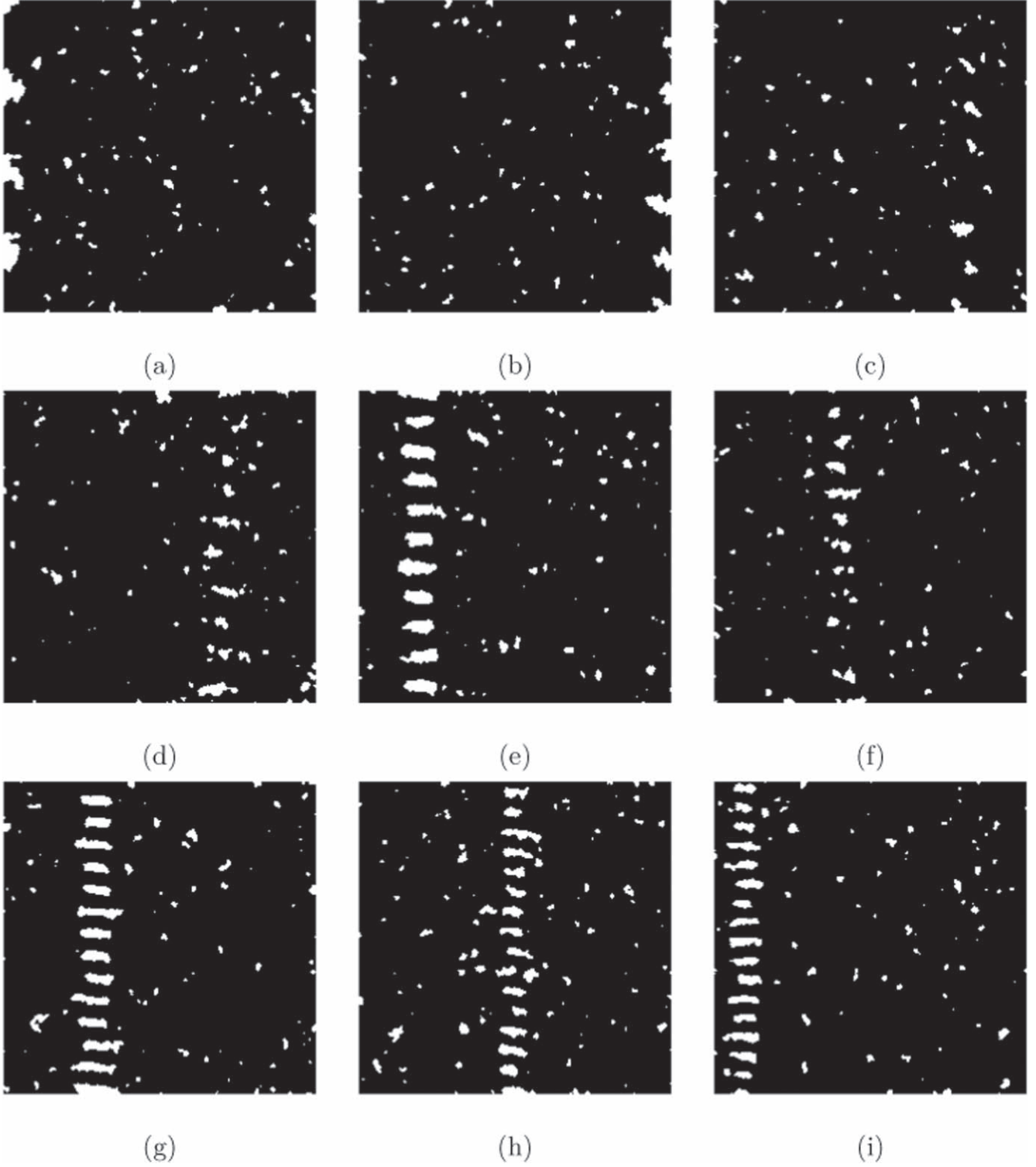


Figure 4. We obtain the segmentation results from the first stage process and can distinguish interferometric fringes explicitly for most images. Through this process, signal and background are distinguished. However, noise can have an influence on location accuracy. Hence, we need to deal with these results through the second stage method.

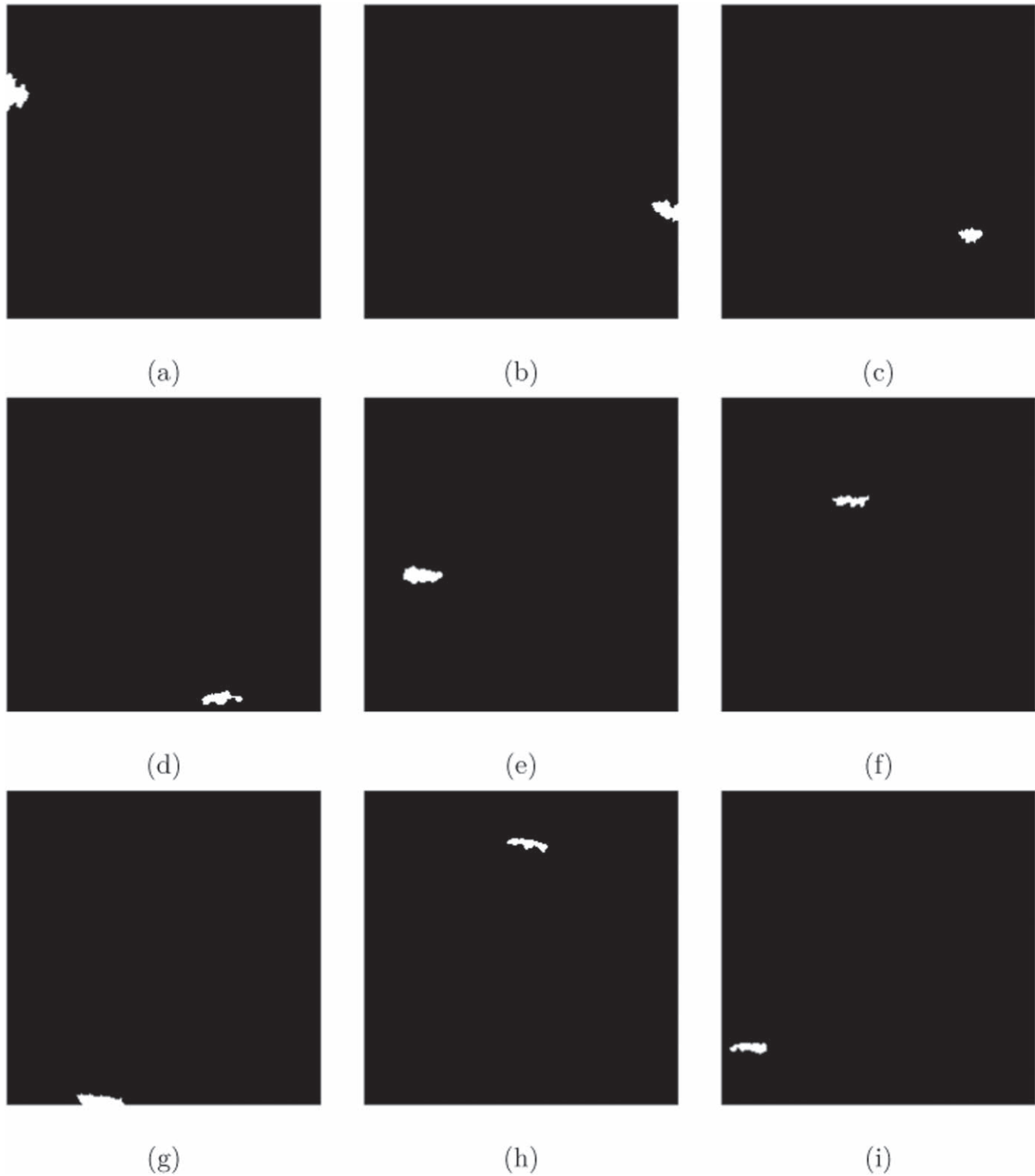


Figure 5. The maximum connection regions are obtained with the seed filling algorithm. Generally speaking, the maximum connection region retains most features of an interferometric fringe. We regard the width of the maximum connection region as the width of the whole interferometric fringe.

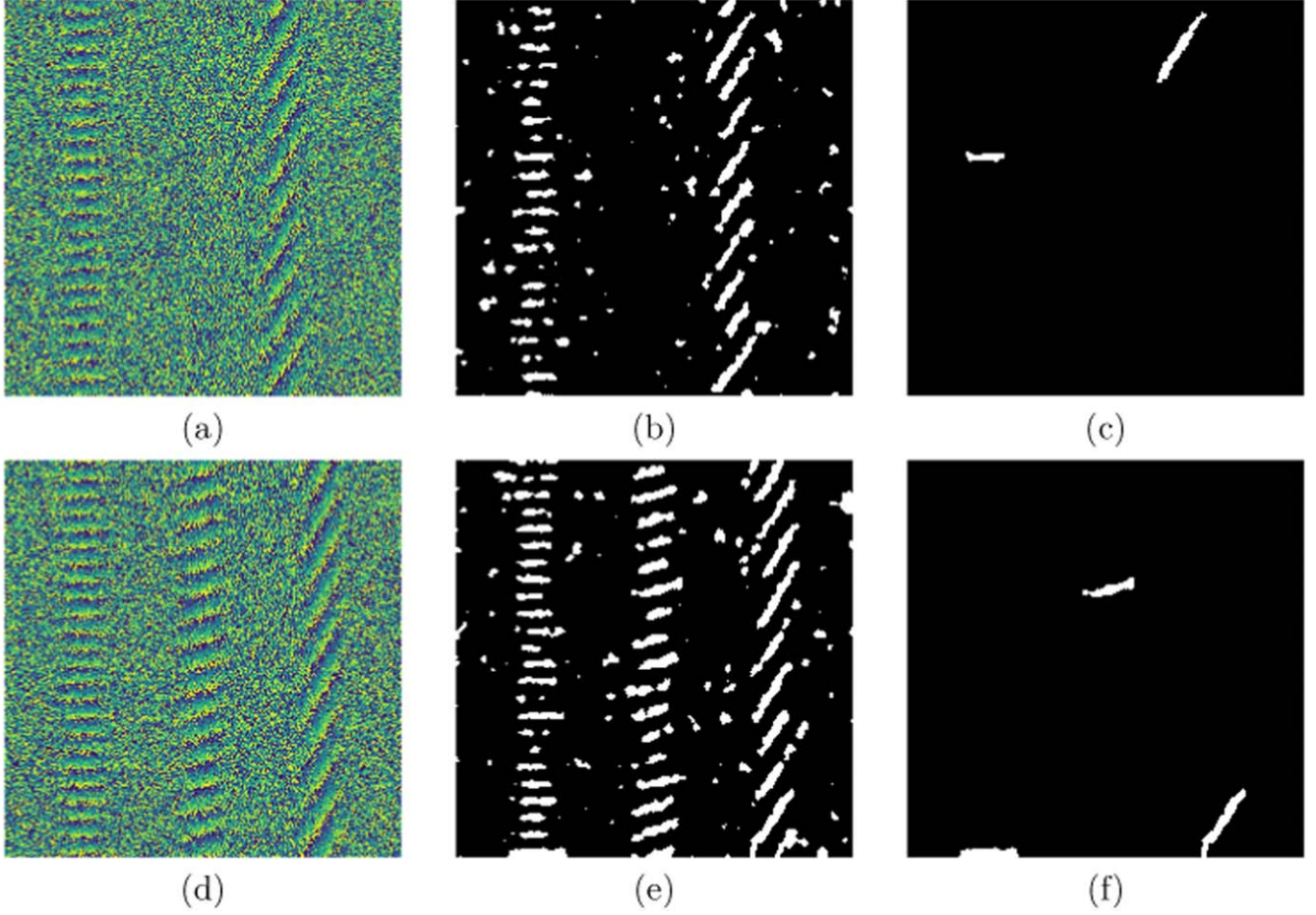


Figure 6. Experimental results with multiple interferometric fringes in the simulation scenario. In this figure, (a) and (d) are original images with 2 and 3 interferometric fringes respectively. (b) and (e) are segmentation results from the first stage. (c) and (f) are the corresponding maximum connection regions of interferometric fringes.

For all visibility images, the origin is located at the bottom left corner, and the horizontal axis is time. The vertical axis represents frequency range of 700–800 MHz. Each image contains only one interferometric fringe. Then, these images with different S/Ns are processed by the KWFLICM clustering algorithm.

The original images with interferometric fringe produced by the simulation program are listed in Figure 3. Some images have low S/N (≤ 0.7), such as Figure 3(c), (d) and (f). Others have high S/N (≥ 1.1), such as Figure 3(e), (g) and (i).

We assign parameters in the KWFLICM algorithm as follows. The exponent factor of membership degree is $m = 2$. In most variants, this index can provide good enough results. We set a threshold $\epsilon = 0.00001$. When $\sum |V^{(b+1)} - V^{(b)}|$ is less than the threshold ϵ , the algorithm is converged. Through adjustment of this parameter, we can balance running time and segmentation accuracy. To obtain a good result, 10^{-5} is commonly used. We assign $winSize = 3$, and a 3×3 window can include local information and the running time of the

algorithm is acceptable. The aim of our method is to distinguish the interferometric fringes of sources from background, which makes signal and noise as separate as possible. We assign number of clusters $c = 2$ to represent signal and background respectively both in the simulated scenario and real scenario. Hence, for these parameters, we do not need more fine-tuning. Because of these almost fixed parameters, our method is convenient to use and operates as under plug and play. Approximately, we can regard it as a parameter free method.

Remark 1 When the first stage process terminates, we can obtain image segmentation results as displayed in Figure 4, in which images have been divided into two regions. We can distinguish clearly the shape of the interferometric fringe in most images. But in some images, the interferometric fringe is not very obvious, which depends on S/N. When S/N is high, the shape of the interferometric fringe is more obvious, therefore, the segmentation result from the first stage algorithm is better. When S/N is lower, the segmentation result is weaker. We can recognize the shape intuitively, such as in

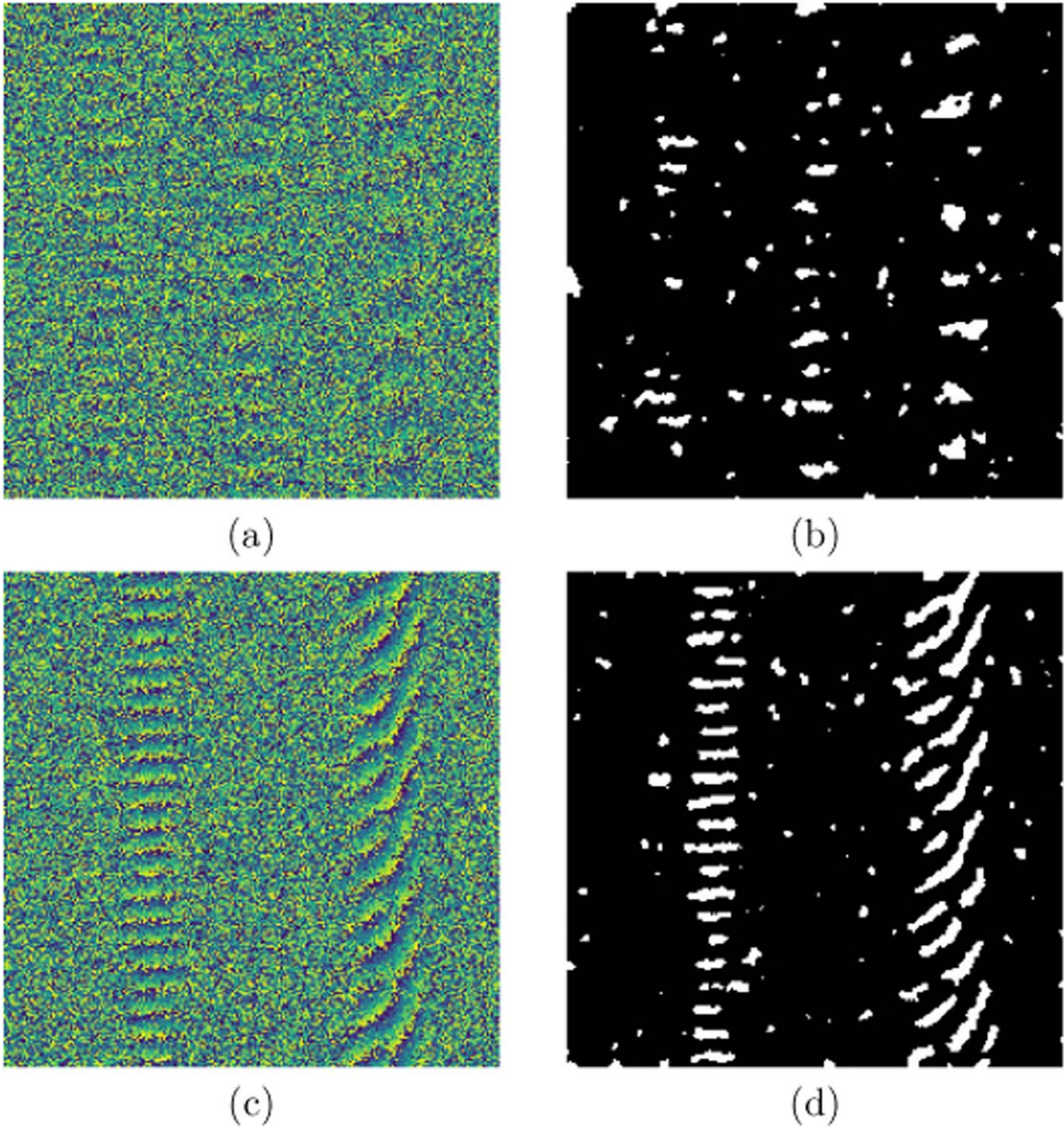


Figure 7. For figure (a), low S/N ($S/N = 0.5$) and non-overlapping fringes can be recognized through our method. For figure (c), overlapping fringes can only be recognized as separate fringes. (b) and (d) are corresponding results.

Figure 4(e) and (g). These results are good, and our method can provide a very useful basis for downstream tasks. Also when S/N is relatively low, the segmentation result still can be used in the next step, such as in Figure 4(h) and (i). It is worth

mentioning that interferometric fringe signals in Figure 4(a) and (b) are located at the left and right borders of the image, respectively. Moreover, the two signals display only a half shape. The algorithm still provides a very good segmentation

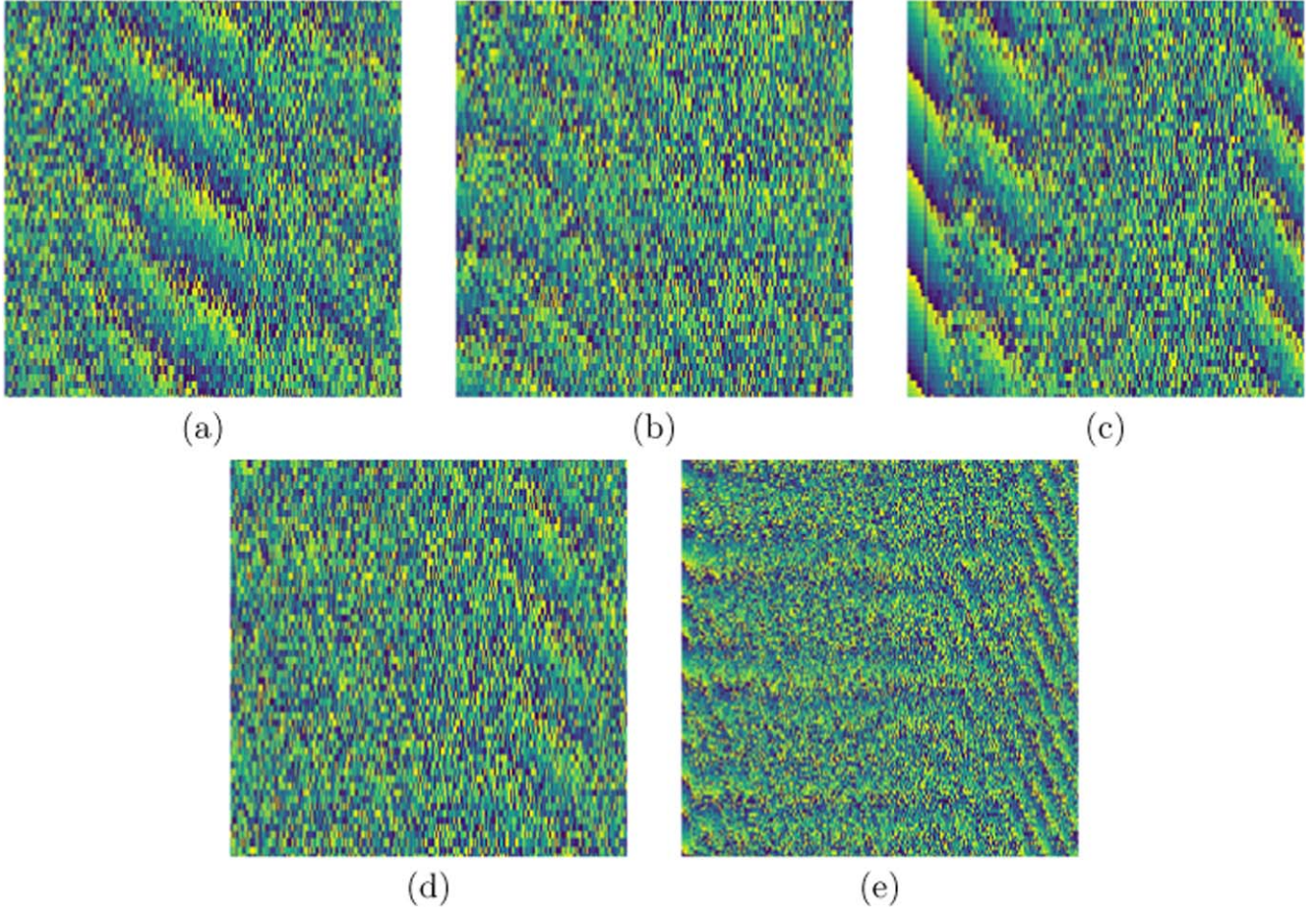


Figure 8. Raw phase images of Tianlai visibility data which are slightly polluted by RFI and cross-coupling. Images include more noises, and more types of fringes are found to be more irregular. Real S/N and other parameters cannot be obtained directly from the real scenario.

result. Furthermore, for an image with the lowest S/N (Figure 4(c)), the algorithm can give a distinct result. For S/N values of images lower than 0.5, we cannot recognize the existence of an interferometric fringe. Hence, we can treat 0.5 as the limiting threshold for this task. For images with higher S/N values ($1.5 \leq S/N \leq 10$), our algorithm can provide very good effectiveness. We can conclude that the first stage algorithm is effective for all S/N visibility data with an interferometric fringe.

After obtaining the segmentation result, we then adopt the seed filling algorithm to search for the maximum connection region. We list the maximum connection region of each image in Figure 5.

Remark 2 Finding the maximum connection region aims at retaining the feature of a fringe as much as possible. Generally speaking, a maximum connection region can be regarded as the most obvious feature of an interferometric fringe. Therefore, an interferometric fringe with the maximum connection region represents an ideal interferometric fringe, which can minimize

the influence from other factors and noise. After finding the maximum connection region, we can distinguish the shape of the interferometric fringe including signal, and to locate the interferometric fringe, we can use the width of the maximum connection region to replace the width of the interferometric fringe signal. In this way, we finally obtain the location of the interferometric fringe.

Remark 3 Table 2 displays the numerical experimental results. In this table, we compute the accuracy through the width of prediction location and real location. Accuracy can be defined as follows:

$$\text{Accuracy}(I) = \frac{S(\text{PreLoc} \cap \text{RealLoc})}{S(\text{PreLoc} \cup \text{RealLoc})},$$

where $\text{Accuracy}(I)$ represents location accuracy of an interferometric fringe from image I . $S(\cdot)$ signifies area. $\text{PreLoc} \cap \text{RealLoc}$ and $\text{PreLoc} \cup \text{RealLoc}$ are intersection and union of prediction location region and real location region respectively.

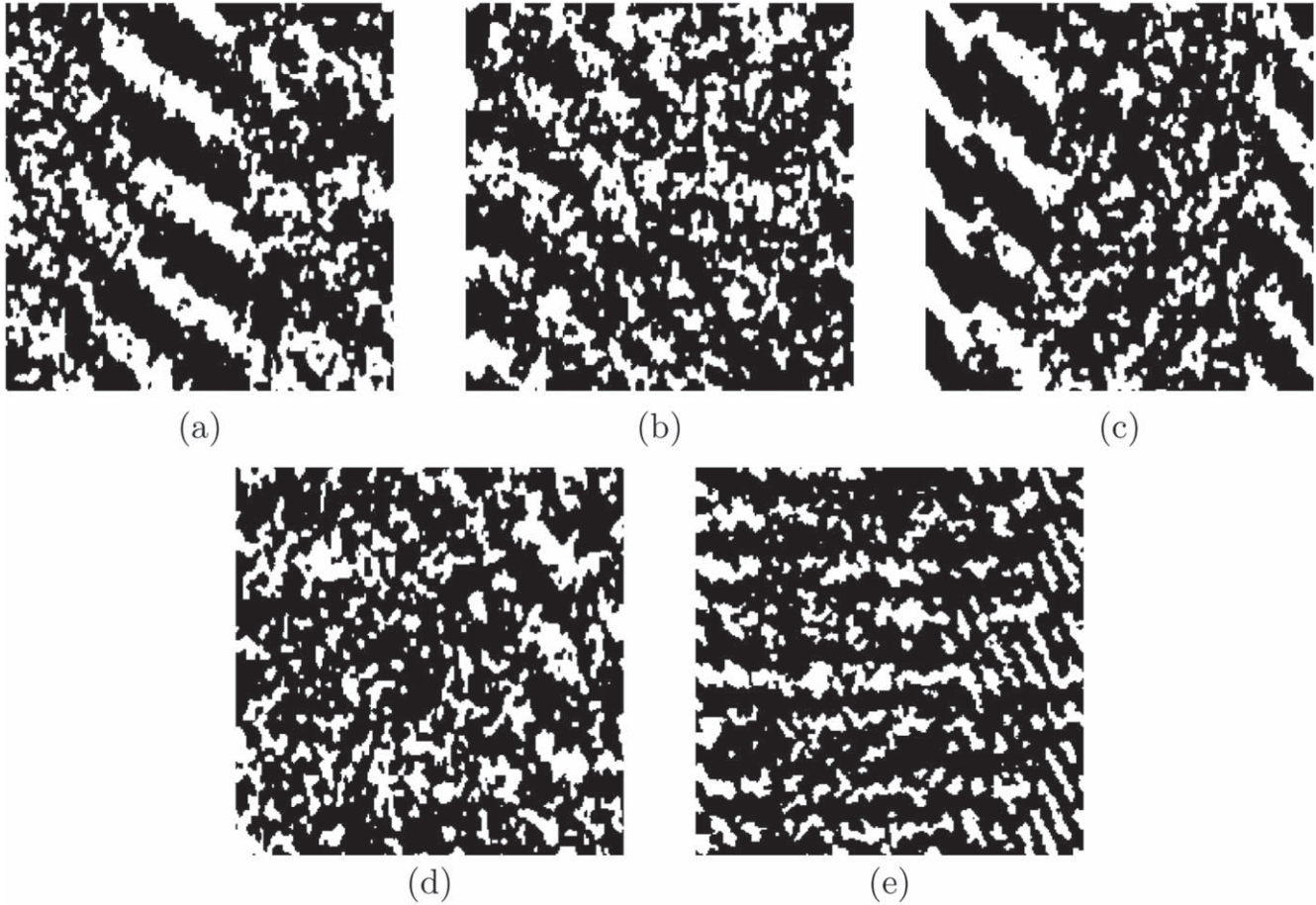


Figure 9. Segmentation results of the real scenario for slightly polluted phase images. Signal and noises are interlaced together and noises are more obvious compared with simulation scenarios.

Table 2
Validation of our Method with Simulation Data

Simulation Image	Prediction Location	Labeled Location	Accuracy(%)	Running Time(s)	Iterations
(a)	(0,0), (15,223)	(0,0), (17,223)	88.2	979	201
(b)	(205,0), (223,223)	(206,0), (223,223)	94.4	553	114
(c)	(169,0), (185,223)	(161,0), (195,223)	47.1	1497	304
(d)	(139,0), (167,223)	(139,0), (173,223)	82.4	2040	415
(e)	(28,0), (55,223)	(28,0), (62,223)	79.4	894	184
(f)	(79,0), (104,223)	(72,0), (106,223)	73.5	5050	977
(g)	(50,0), (83,223)	(50,0), (84,223)	97.1	1004	202
(h)	(102,0), (130,223)	(94,0), (128,223)	76.5	1520	312
(i)	(6,0), (31,223)	(5,0), (39,223)	73.5	1160	233

In Table 2, for those images with lower S/N, the effect is weak, such as in Figure 5(c), (d), and (f). For Figure 5(b) and (g) with higher S/N, the algorithm will provide better results. These numerical experiments accord with our intuition. Generally, because every simulation image has the same width

and height, the running time of every iteration is comparable. Iterations are related to the characteristics of the interferometric fringe and noise distribution in an image.

In Table 2, some accuracies are relatively low in numeric values, such as (c) and (f), which are the lowest S/N images.

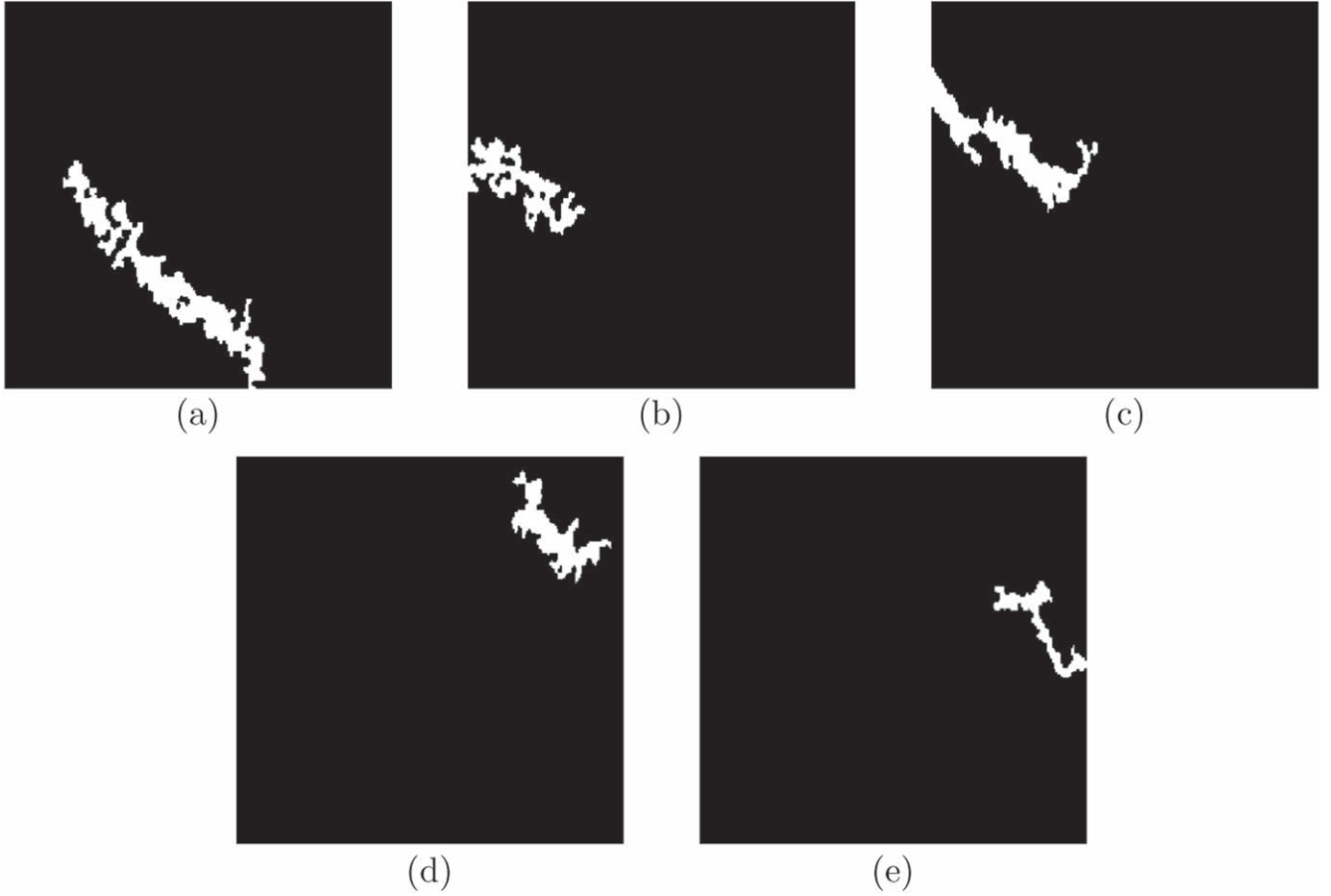


Figure 10. Maximum connection regions of images in the real scenario for slightly polluted phase images. The results are different from the simulation scenario. The shapes of these regions are more irregular.

Table 3
Validation of our Method with Real Data

Real Image	Labeled Location	Detected Location	Accuracy(%)	Running Time(s)	Iterations
(a)	(40,0), (145,223)	(34,0), (150,223)	90.5	5214	940
(b)	(0,0), (70,223)	(0,0), (66,223)	94.4	2265	407
(c)	(0,0), (78,223)	(0,0), (95,223)	82.1	2757	485
(d)	(152,0), (208,223)	(159,0), (216,223)	76.6	2257	405
(e)	(178,0), (223,223)	(170,0), (223,223)	84.9	8494	1459

Though our method can only provide limited accuracy, we can still give the effective location of the interference fringe. It is also worth mentioning that the criterion of accuracy is the ratio of the intersection area to the union area of prediction location and labeled location. Since even a few fixels' mismatch in the time axis can lead to a several percentage drop in area ratio, this accuracy criterion can easily magnify any small mismatch. Hence, for images with very weak S/N, the accuracy of 47% is still acceptable.

4.2.2. Multiple Interferometric Fringes

In drift scan observations, if multiple sources are nearby in right ascension (R.A.), there can be multiple fringes in our randomly chosen time interval. In the simulation scenario, we also consider such case so as to explore the generalization ability of our method. We give segmentation and use the eraser algorithm in Section 3 to obtain interferometric fringe locations in Figure 6. This experiment shows the effectiveness of our

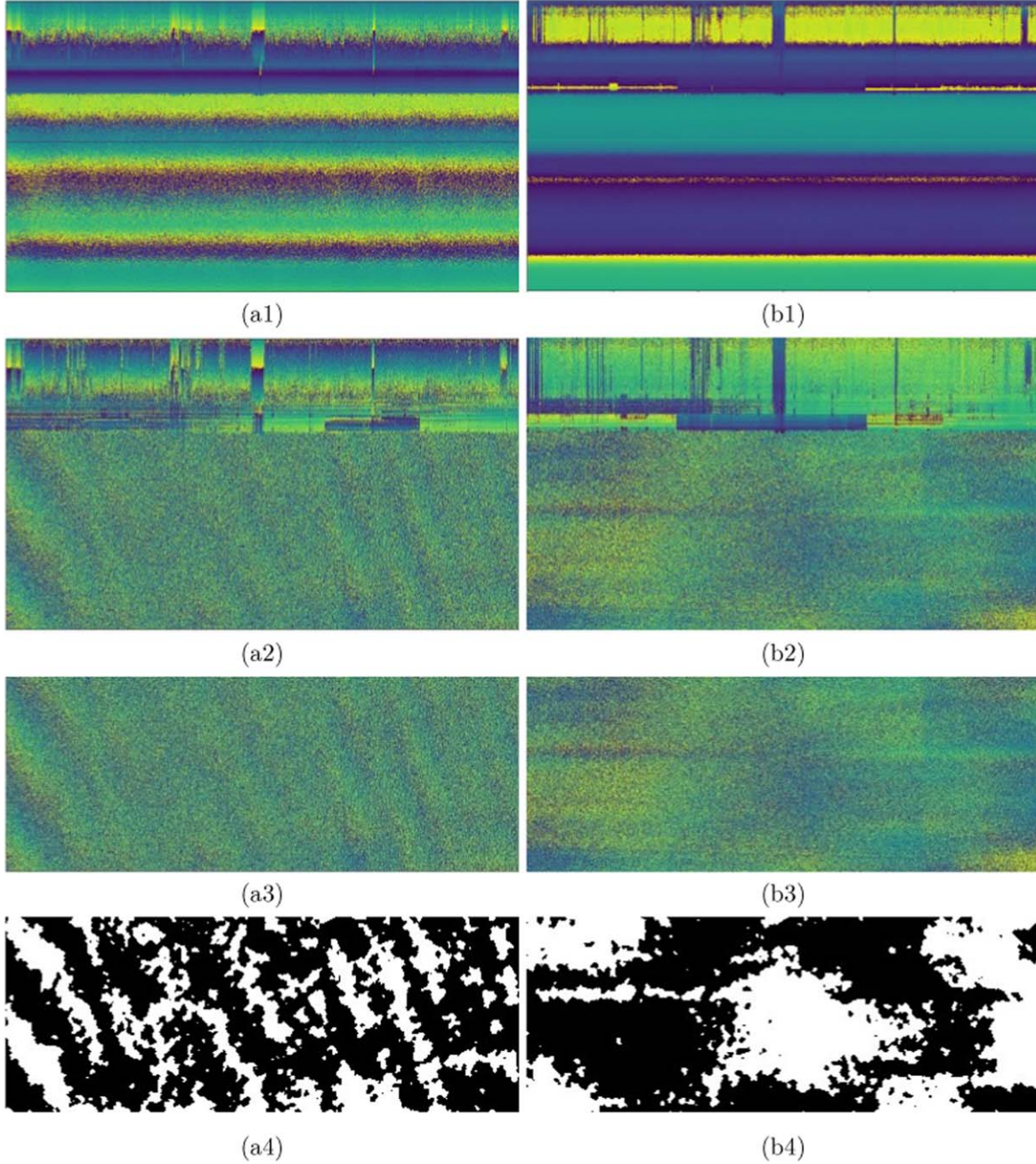


Figure 11. Results of serious RFI polluted images processed by our method. (a1) and (b1) are raw phase images. Serious narrow band RFI intermittently appears in the frequency range of 760–788 MHz (the top part of Figure (a1) and (b1)). Strong cross-coupling effects also exist, which are stable horizontal phases throughout the whole time interval. (a2) and (b2) are phases after cross-couplings are removed. (a3) and (b3) are phases after the RFI frequency range are cut-off. (a4) and (b4) are the segmentation results.

method in a multiple fringe scenario. When S/N values of multiple interferometric fringes are close, our method can give explicit results. However, when their S/N values have big differences, some interferometric fringes with higher S/N can suppress those with lower S/N. Some interferometric fringes with low S/N may be discarded as noise.

For multiple fringe detection, the maximum number of fringes that can be effectively recognized depends on the complexity of images. For non-overlapping and close S/N fringes, they can all be recognized, so the maximum number depends on the sources' number. For overlapping fringes, the maximum number of recognized fringes of our method is

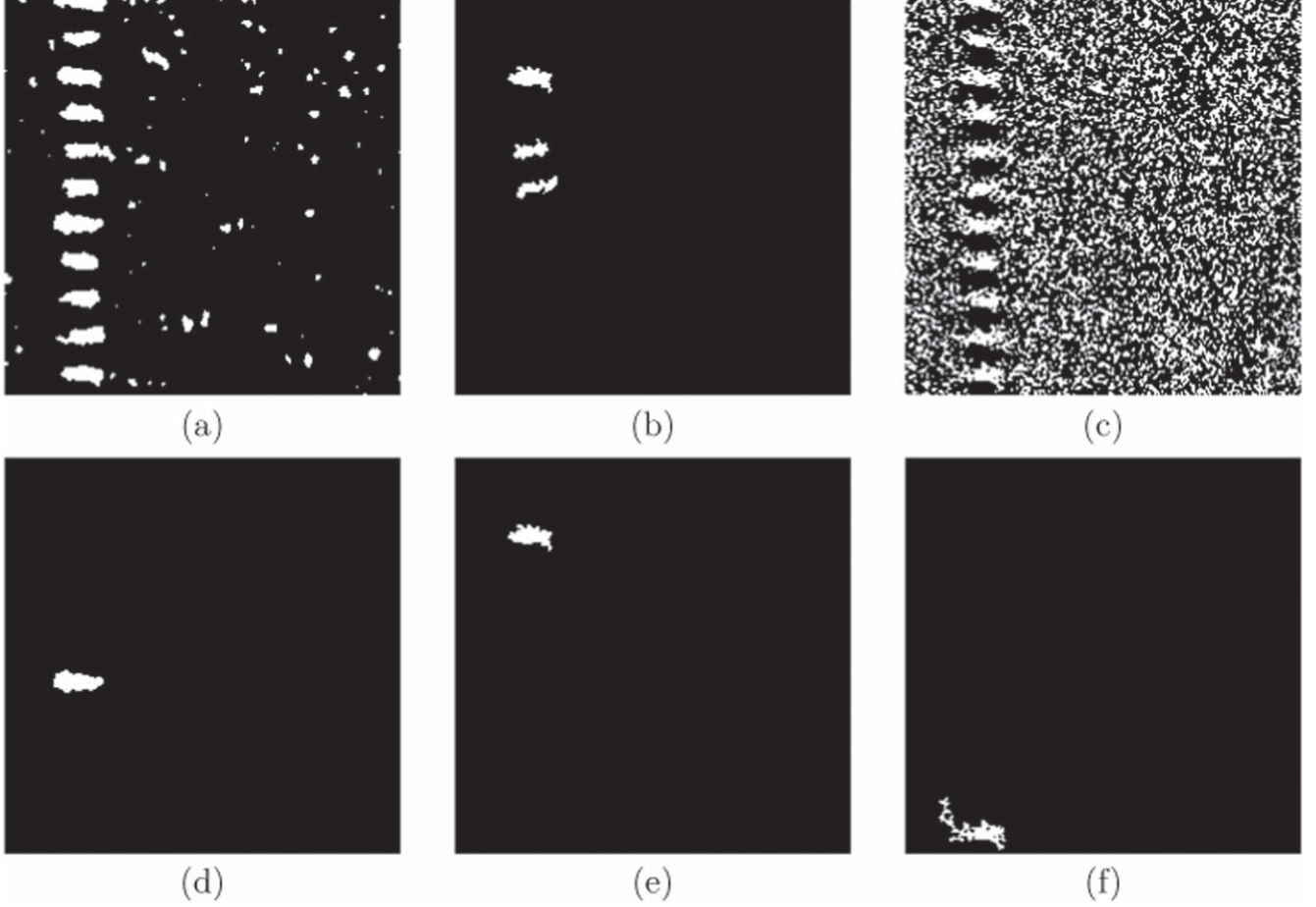


Figure 12. Comparison of experiments between our method and two state of the art methods. We integrate these three unsupervised segmentation methods with searching for the maximum connection region. Results (a) and (d) are from our method, (b) and (e) are from DFKM (Zhang et al. 2020), and (c) and (f) are from an unsupervised image segmentation by backpropagation (Kanezaki 2018). The accuracies of these three methods are 79.4%, 70.6% and 65.1%.

restricted by overlapping fringes. When multiple fringes overlap, some fringes will combine into one fringe, and the number of maximum recognized fringes will be less than the number of fringes. To illustrate this result, we show the experimental results in Figure 7. For image (a), three fringes have low S/N of 0.5. Our method can identify the shape of three fringes, and the number of maximum recognized fringes is 3. For image (c), two single fringes are overlapped because of their locations in the image. Overlapped fringes will be regarded as one fringe, and the number of maximum recognized fringes will be 2. In this case, although we cannot obtain the width of the interferometric fringes directly, we can provide a basis for further processing. Based on this result, we can consider developing a method for unmixing overlapped interferometric fringes in the special downstream physical parameter estimation task for this case. On the other hand, if not all fringes are overlapped, for non-overlapped fringes, our method still gives explicit results.

4.3. Real Scenario

In this experiment, we use real images with an interferometric fringe from the Tianlai Cylinder telescope array to validate our method effectiveness further. Compared to the simulation scenario, the real observed data will inevitably have more noises, such as radio frequency interference (RFI) from the surrounding environment or cross-coupling effect for densely distributed radio arrays. These noises are either slight or serious, which will have different impacts on the fringe recognition. We will discuss the two cases separately.

4.3.1. Slightly Polluted Images

In Figure 8, we list real images with an interferometric fringe from different S/Ns. Some of them have obvious fringes, such as Figure 8(a), (c), and (e), and others have weak fringes, such as Figure 8(b) and (d). Compared to images in the simulation scenarios, noises are more obvious, which will have an influence on the location effectiveness. However, these images

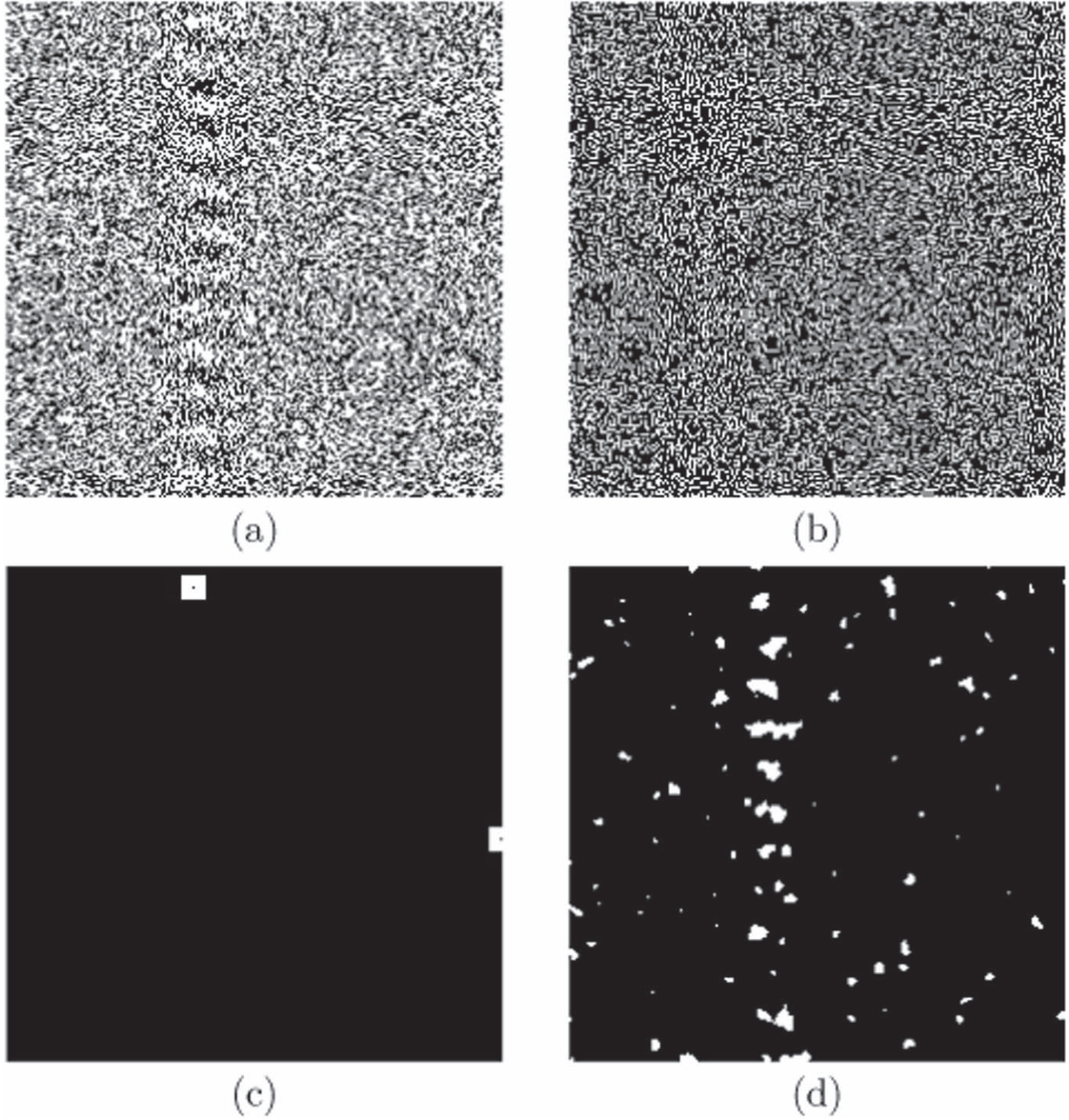


Figure 13. Comparison experiments between our first stage method and classical image segmentation methods. Results show our method can give a clearer shape of interferometric fringes.

are still only slightly polluted by RFI and cross-coupling. For this kind of image, we can directly apply our method in Section 3 to these images. After the first stage process, we obtain segmentation results of real scenario images in Figure 9.

Further, we adopt the second stage algorithm to locate the interferometric fringe, and these results are listed in Figure 10.

Remark 4 We can distinguish segmentation results in Figure 9, which are different from those in Figure 4. After

the first stage process, we can distinguish an interferometric fringe obviously in Figure 4. However, in Figure 9 it is not easy to distinguish an interferometric fringe because of the influence of more factors and noises. But after the maximum connection region algorithm is executed, the location of an interferometric fringe can be detected successfully, see Figure 10. Therefore, not only in the simulation scenario but also in the real scenario, for slightly polluted images, our method shows good effectiveness. It does not need any pre-processing and exhibits robustness (numeric results in Table 3). But compared to the simulation scenario, our method in the real scenario consumes more running time and iterations, which illustrates that images in the real scenario have more complexity.

Remark 5 For both the simulation scenario and real scenario, we can distinguish locating effectiveness for interferometric fringes intuitively. In the simulation scenario, we give a definition of accuracy, and this criterion shows a numeric difference in different images. For interferometric fringes with different S/Ns, even with some special conditions, our method can provide good location results ($\geq 70\%$). In real scenarios, the lack of labels is a problem, so we employ experts to complete labeling real images to obtain the ideal width of the interferometric fringe. Further, our algorithm gives relatively strong location accuracy ($\geq 80\%$).

4.3.2. Seriously Polluted Images

In actual data processing, the RFI can sometimes become very strong. It will generate fake fringes in the phase images which would probably affect the detection of weak signals. Besides, the cross-coupling effect for short baselines of a radio array will behave stronger than longer baselines. Hence, we add experiments about data polluted by strong RFI and the cross-coupling effect. The results are displayed in Figure 11. Figures (a1) and (b1) is the raw phase images. We find that the whole images are full of very strong cross-coupling effect, which are those stable horizontal phases throughout the whole time interval. The cross-couplings generate much stronger fringes, so the interference fringes are buried by the cross-couplings. However, the cross-couplings are very stable in the time interval, so we can remove the time averaged phase to get rid of the cross-coupling effect. These results are shown in Figures (a2) and (b2). Moreover, in the top part of Figures (a1) and (b1), a serious narrow band RFI intermittently appears in the frequency range of about 760–788 MHz. This level of RFI will totally break the source's fringe continuity in the frequency direction, especially for weak sources. To detect the location of a fringe, this polluted frequency range must be cut off before the data processing is complete. The results are displayed in Figures (a3) and (b3). Finally, we use the proposed method to complete interferometric fringe recognition. The results are depicted in Figures (a4) and (b4).

From these results, we can know that for images affected by strong RFI and cross-couplings, a pre-process—narrow band RFI and cross-coupling removal—can help reduce the influence of noise. Our method can then handle images containing fringes easily and directly. Therefore, for seriously polluted images, such as narrow band RFI and cross-couplings, a pre-process is needed before carrying out fringe detection using our method. Then the fringe detection issue becomes a problem, the same as the slightly polluted case. Considering that many images are polluted by a different variety in interference factors in the real scenario, in most cases, pre-processing of the data is required before the actual application of the proposed method.

4.4. Comparison with State of the Art Methods

In this section, we make comparisons with state of the art methods to illustrate adaptability and validation of our method. Some unsupervised methods have shown good performance in other fields of image processing, such as the deep fuzzy K-means (DFKM) algorithm (Zhang et al. 2020) and unsupervised image segmentation by propagation (unprop) (Kanezaki 2018). In this subsection, we complete comparison experiments in simulated images between our method and these methods. We apply these three methods in the simulated image of Figure 3(e) and obtain location results. Results from different methods and their corresponding maximum connection regions are listed in Figure 12. Further, we can know that the widths of location results of three methods are [(28,0), (55,223)] (Our method), [(30,0), (54,223)] (DFKM) and [(19,0), (56,223)] (unprop), respectively. Since the labels are (28,0), (62,223), we can compute the location accuracy of the three methods to be 79.4%, 70.6% and 65.1%, respectively. Our method shows an obvious numeric advantage.

4.5. Comparison with Classical Image Segmentation Methods

In the proposed method, image segmentation is an important step for detection of interferometric fringes. Some classical image segmentation methods play a critical role aiming at a different variety of images. To illustrate the effectiveness of our first stage method, we further compare with classical image segmentation methods in Figure 3(f). Results are shown in Figure 13. (a) represents segmentation result from the Otsu method (Otsu 1979), which is representative of a threshold based image segmentation. (b) is the result from the Canny method (Canny 1986), which is an edge based method. (c) is the result from the Watershed method (Vincent & Soille 1991), which is a region based method. (d) is the result from our method. From these results, we can know that our method can retain more shape features of interferometric fringes. For the Otsu method and Canny method, these methods only provide limited results, and we hardly recognize the shape of

interferometric fringes. For the Watershed method, it can provide results for a rectangular region. However, these results hardly display shape features, which introduces confusion in the next step in processing. Hence, comparing with some classical image segmentation methods, our method can provide a good basis for subsequent image processing.

5. Conclusion

In the source detection problem for radio telescope arrays such as the Tianlai project, how to effectively locate interferometric fringes in massive visibility data with a high efficiency remains a problem. In this paper, we investigate the characteristics of an interferometric fringe in different scenarios. Furthermore, we propose a hierarchical method to locate the interferometric fringe. In the first stage, we regard this task as image segmentation, and introduce an unsupervised clustering algorithm. In the second stage, we use a seed filling algorithm to find the maximum connection region. Then we can regard the location of the maximum connection region as the location of the source's interferometric fringe. Finally, we validate our method in real scenarios which are slightly and seriously polluted by RFI and cross-couplings, and make comparisons with state of the art methods to illustrate the effectiveness.

In the future, we will focus on the differences between interferometric fringe signals and noises, and extract more useful features to enhance the location accuracy, and propose more effective methods for fringe location, especially to remove RFI and cross-coupling automatically. On the other hand, we will propose physical parameter estimation with unsupervised learning. Considering the running speed, we will apply our method in a GPU environment to make the algorithm run in real time. We hope that it will enrich the community and provide more novel data processing methods for weak source detection.

Acknowledgments

The authors are grateful to the chief of the Cosmology Dark Matter and Dark Energy Group at National Astronomical Observatories, Chinese Academy of Sciences, Professor Xuelei Chen, and chief of the Comprehensive Astrophysics Research Group, Professor Wenwu Tian at National Astronomical Observatories, Chinese Academy of Sciences, and Dr. Dan Wu and their teams, for their splendid research work and suggestions. The authors are grateful to the anonymous

reviewers and Editor-in-Chief Professor Liang Gao for their excellent comments and valuable suggestions that helped us improve this paper. The authors are grateful to Professor Dan Hu at the University of North Carolina at Chapel Hill. This study is supported by the National Natural Science Foundation of China (NSFC, grant Nos. 42172323 and 12371454). The authors declare there is no conflict of interest.

References

- Awni, Y., Pranav, R., Masoumeh, H., Geoffrey, H., & Codie, B. 2019, *Nature Medicine*, 25, 65
- Canny, J. 1986, *ITPAM*, 8, 679
- Cao, X., Chen, B., & He, W. 2022, *ITIM*, 71, 1
- Cavaglia, M., Staats, K., & Gill, T. 2018, *CCoPh*, 25, 1
- Chen, X. 2011, *SSPMA*, 5304, 43
- Chen, X. 2012, *Int. J. Modern Phys.: Conf. Ser.*, 12, 256
- Dunn, J. C. 1973, *Journal of Cybernetics*, 3, 32
- Everett, S., Yanny, B., Kuropatkin, N., et al. 2022, *ApJS*, 258, 15
- Fong, S., & Narasimhan, S. 2022, *ITIM*, 71, 1
- Fu, Y., Fan, J., Xing, S., et al. 2022, *ITIM*, 71, 1
- Gheller, C., Vazza, F., & Bonafede, A. 2018, *MNRAS*, 480, 3749
- Gong, M., Liang, Y., Shi, J., Ma, W., & Ma, J. 2013, *ITIP*, 22, 573
- Hathaway, R. J., & Bezdek, J. C. 2000, *IEEE Transactions on Fuzzy Systems*, 8, 576
- He, J., Ouyang, M., Chen, Z., Chen, D., & Liu, S. 2022, *ITIM*, 71, 1
- He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conf. on Computer Vision and Pattern Recognition (Las Vegas, NV: IEEE), 770
- Kanezaki, A. 2018, in *Icassp IEEE Int. Conf. on Acoustics* (Calgary, AB: IEEE),
- Korytov, D., Hearin, A., Kovacs, E., et al. 2019, *ApJS*, 245, 26
- Krinidis, S., & Chatzis, V. 2010, *ITIP*, 19, 1328
- Kuang, J., Xu, G., Tao, T., & Wu, Q. 2022, *ITIM*, 71, 1
- Li, J., Zuo, S., Wu, F., et al. 2020, *SCPMA*, 63, 129862
- Lin, A., Chen, B., Xu, J., et al. 2022, *ITIM*, 71, 1
- Liu, W., Zhu, M., Dai, C., He, D., & Yu, X. 2019a, *MNRAS*, 483, 4774
- Liu, W., Zhu, M., Dai, C., et al. 2019b, *RAA*, 19, 1
- Memon, K., & Lee, D. 2018, *Fuzzy Sets and Systems*, 340, 91
- Otsu, N. 1979, *ITSMC*, 9, 62
- Peng, C., Liu, Y., Gui, W., Tang, Z., & Chen, Q. 2022, *ITIM*, 71, 1
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016, in *IEEE Conf. on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 779
- Shelhamer, E., Long, J., & Darrell, T. 2017, *ITPAM*, 39, 640
- Tanoglidis, D., Drlica-Wagner, A., Wei, K., et al. 2021, *ApJS*, 252, 18
- Thompson, A., Moran, J., Swenson, G., & Aurass, H. 1991, *Interferometry and Synthesis in Radio Astronomy* (New York: Wiley)
- Vincent, L., & Soille, P. 1991, *ITPAM*, 13, 583
- Wang, H., Zhu, W., Guo, P., & Li, D. 2019, *SCPMA*, 62, 61
- Xu, Z., Yu, F., Chang, S., & Wang, S. 2015, arXiv:1503.00591
- Yan, R., Ma, R., Liu, W., et al. 2022, *DSP*, 129, 103663
- Zadeh, L. A. 1965, *InfCo*, 8, 338
- Zeng, W., Ma, R., Yin, Q., Zheng, X., & Xu, Z. 2020, *JIFS*, 39, 1
- Zeng, X., Wei, S., Shi, J., & Zhang, X. 2021, *ITIM*, 70, 1
- Zhang, R., Li, X., Zhang, H., & Nie, F. 2020, *IEEE Transactions on Fuzzy Systems*, 28, 2814
- Zhu, Y., Liang, X., Wang, T., Xie, J., & Yang, J. 2022, *ITIM*, 71, 1
- Zuo, S., Li, J., Li, Y., et al. 2021, *A&C*, 34, 100439